The Dirichlet-Multinomial Log-Likelihood Function

Djallel Bouneffouf

HAL Id: hal-03008622
https://hal.science/hal-03008622

Submitted on 16 Nov 2020

The Dirichlet-Multinomial Log-Likelihood Function

Djallel Bouneffouf[1]

IBM Research

Abstract. Dirichlet-multinomial (DMN) distribution is commonly used to model over-dispersion in count data. Precise and fast numerical computation of the DMN log-likelihood function is important for performing statistical inference using this distribution, and remains a challenge. To address this, we use mathematical properties of the gamma function to derive a closed form expression for the DMN log-likelihood function. Compared to existing methods, calculation of the closed form has a lower computational complexity, hence is much faster without comprimising computational accuracy.

1 Introduction

The uses of the binomial and Multinomial (MN) distributions in statistical modelling are very important, with a huge variety of applications in bioinformatics [1]. The modeling of over-dispersion of the MN distribution has been solved by extending the MN distribution to the Dirichlet-multinomial (DMN) distribution [2].

The computation of the DMN log-likelihood function by existing methods leads to non accurate results [3]. Recently, an alternative formulation done in [3] solve this issue, but it still leads to long runtimes that make it inappropriate for large count data. We have developed a new method for computation of the DMN log-likelihood to solve the accuracy problem without incurring long runtimes.

The new approach is mainly based on the definition of the Dirichlet distribution according to the Gamma function. Our numerical experiments show that this new method improves the runtime in high-count data situations that are common in deep sequencing data, and then increases the usefulness of the DMN distribution to model many bioinformatics problems.

2 Previous work

Due to its demonstrated effectiveness and its mathematical convenience, the DMN distribution is widely applied into different domains, such as multiple sequence alignment [1] in bioinformatics, natural language processing [4] [5], finance [6] and online learning [7–31] .

Likelihood functions play a important role in statistical, where it is generally used for parameter estimation and interval estimation [32]. Unfortunately, numerical computation of the DMN log-likelihood function by conventional methods results in instability in the neighborhood of $\phi = 0$, where the parameter ϕ characterizes how different a DMN distribution is from the corresponding MN distribution with the same category

probabilities, and gives the DMN distribution the ability to capture variation that cannot be done by the MN distribution. Hence, the functions implementing of the DMN done in [33], [34] or [35], return NaN when $\phi = 0$.

To overcome this instability of accuracy on computing the log-likelihood, authors in [3] derived a new approximation of the DMN log-likelihood function based on Bernoulli polynomials. Using a mesh algorithm, they computed the log-likelihood for any parameters in the domains of the log-likelihood function. Comparing with the existing methods, the proposed method is more accurate and is faster.

In this paper we propose a much faster method for the computation of the DMN log-likelihood function and we demonstrate it's efficiency though simulation.

3 DMN DISTRIBUTION

Following the notation done in [3], the probability mass function (PMF) of the K categories MN distribution of N-independent trials is given by

$$f_{MN}(x; N, p) = \frac{N!}{\prod_{k=1}^{K} x_k!} \prod_{k=1}^{K} p_k^{x_k} \tag{1}$$

where $N!$ denotes the factorial of a non-negative integer N; the observations $x = (x_1, ..., x_k)$, satisfying $\sum_{k=1}^{K} x_k = N$, are non-negative integers; and $p = (p_1, ..., p_k)$, satisfying $\sum_{k=1}^{K} p_k = 1$, are the probabilities that these K categories occur.

The DMN distribution is generated with the probabilities p following a prior distribution (of the positive parameters $\alpha = (\alpha_1, ..., \alpha_K)$) conjugate to the PMF $f_{MN}(x; N, p)$ [36] as follows:

$$f_{Dir}(p; \alpha) \propto \prod_{k=1}^{K} p_k^{\alpha_i - 1} \tag{2}$$

The normalized form of the Dirichlet distribution is as follows:

$$f_{Dir}(p; \alpha) = \frac{N!}{\prod_{k=1}^{K} x_k!} \prod_{k=1}^{K} p_k^{x_k} \tag{3}$$

where $\Gamma(x)$ is the gamma function and $A = \sum_{i=1}^{K} \alpha_i$.

By taking the integral of the product of the MN likelihood (1) the Dirichlet prior (2) and with respect to the probabilities p, the PMF of the DMN distribution is derived [2],

$$f_{Dir}(x : N; \alpha) = \frac{N!}{\prod_{k=1}^{K} x_k!} \frac{\Gamma(A)}{\Gamma(A + N)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + x_k)}{\Gamma(\alpha_k)} \tag{4}$$

where, same as the MN distribution, $x = (x_1, ..., x_K)$ are non-negative integers, satisfying $N = \sum_{k=1}^{K} k$. The first term on the right side of (3) does not depend on the parameter α. For common uses of the likelihood function in the maximum-likelihood

estimation, we are only interested in the last two terms of the DMN likelihood function in (3), then:

$$L(\alpha : x) = \frac{\Gamma(A)}{\Gamma(A+N)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_k + x_k)}{\Gamma(\alpha_k)} \qquad (5)$$

4 PROPOSITION

To compute the log-likelihood of the DMN we are using the definition of the Dirichlet-distribution. According to the definition of the Γ function we have,

$$\Gamma(A) = (A - 1)\Gamma(A - 1) \qquad (6)$$

then

$$\frac{\Gamma(A)}{\Gamma(A+N)} = (\prod_{i=A}^{A+N-1} i)^{-1} \qquad (7)$$

and

$$\frac{\Gamma(\alpha_k + x_k)}{\Gamma(\alpha_k)} = \prod_{i=\alpha_k}^{\alpha_k + x_k - 1} i \qquad (8)$$

then

$$L(\alpha, x) = \frac{\prod_{k=1}^{K} \prod_{j=\alpha_k}^{\alpha_k + x_k} j}{\prod_{k=A}^{A+N-1} i} \qquad (9)$$

By taking the logarithm of both sizes of (9), we get the log-likelihood function

$$\log(L(\alpha, x)) = \sum_{k=1}^{K} \sum_{j=\alpha_k}^{\alpha_k + x_k - 1} log(j) - \sum_{k=A}^{A+N-1} log(i) \qquad (10)$$

We call this new method of computing the Dirichlet-multinomial distribution "Exact-DMN".

4.1 Complexity

Table 1 gives the complexity of each methods.

We observe from the the table that $ExactDMN$ implementation is less time consuming then the $VGAM$ and Mesh implementations, with $m > N$ and m .

Table 1: The DMN method's complexity

Methods	Complexity
VGAM [37]	$O(A + N - 1) + O(A - 1)+$ $\sum_{k=1}^{K} O(\alpha_k + x_k + \alpha_{k-2})$
Mesh [3]	$O(m) + O(\sum_{k=1}^{K}(m + x_k))$
ExactDMN	$O(\sum_{k=1}^{K}(x_k - 1)) + O(N - 1)$

5   Experiments

We implemented the algorithm for computing the DMN log-likelihood in java. In this section, we demonstrate the accuracy and run-time of the proposed algorithm. All experiments were run on a Linux machine with a 4-core Intel Xeon CPUs E5630@ 3.53GHz.

We compute the error of the ExactDMN algorithm by comparing its results with the results of an implementation of (4) in [38], which can achieve high accuracy [3].

Figure 1 compares the error of the DMN methods where the parameters are $x = n(1, 1, 1, 1)$, $\phi = 1/200$ and $p = (.1, .2, .3, .4)$. We can see the ExactDMN algorithm is more accurate then VGAM and has the same accuracy as mesh algorithm.
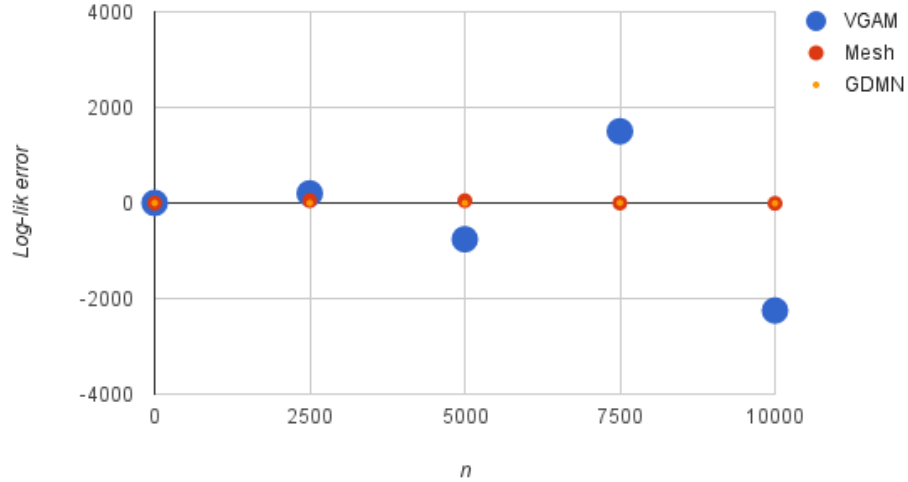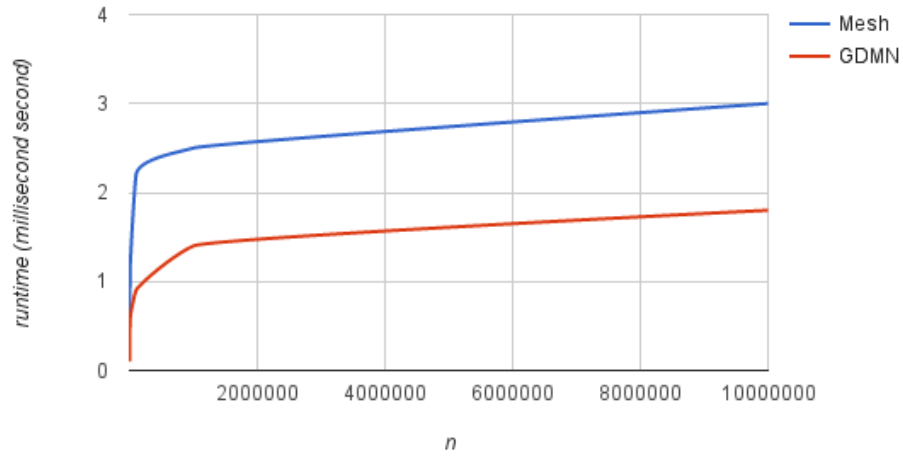
Fig. 1: DMN Method's error with log-lik error (10^{-14})

Figure 2 shows that the runtime of the DMN methods with the following parameters, $x = n(1, 2, 3)$, $p = (1/6, 1/3, 1/2)$ and $\phi = 1/60$, where each plot represents 100 DMN log-likelihood evaluations. We can see from the Figure 2 that the ExactDMN

Fig. 2: DMN Method's runtime

algorithm is much faster than the mesh algorithm for the DMN log-likelihood computation. DMNG algorithm increases more slowly as the counts n increase than the method in mesh.

6   Conclusion

Over-dispersion is important and needs to be accommodated in modeling count data. To handle over-dispersion in MN data, the DMN distribution is commonly used, but its computation is still a challenge. Mainly based on the definition of the Dirichlet distribution, our method solves the runtime challenge without undergoing errors.

References

1. Brown, M., Hughey, R., Krogh, A., Mian, I.S., Sjölander, K., Haussler, D.: Using dirichlet mixture priors to derive hidden markov models for protein families. In: PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS FOR MOLECULAR BIOLOGY, AAAI Press (1993) 47–55
2. Mosimann, J.E.: On the compound multinomial distribution, the multivariate distribution, and correlations among proportions. Biometrika **49**(1/2) (1962) pp. 65–82

3. Yu, P., Shaw, C.A.: An efficient algorithm for accurate computation of the dirichlet-multinomial log-likelihood function. Bioinformatics (2014) btu079

4. Mimno, D., McCallum, A.: Topic models conditioned on arbitrary features with dirichlet-multinomial regression. arXiv preprint arXiv:1206.3278 (2012)

5. MacKay, D.J., Peto, L.C.B.: A hierarchical dirichlet language model. Natural language engineering **1**(03) (1995) 289–308

6. Madsen, R.E., Kauchak, D., Elkan, C.: Modeling word burstiness using the dirichlet distribution. In: Proceedings of the 22nd international conference on Machine learning, ACM (2005) 545–552

7. Bouneffouf, D., Bouzeghoub, A., Gançarski, A.L.: Risk-aware recommender systems. In Lee, M., Hirose, A., Hou, Z., Kil, R.M., eds.: Neural Information Processing - 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part I. Volume 8226 of Lecture Notes in Computer Science., Springer (2013) 57–65

8. Choromanska, A., Cowen, B., Kumaravel, S., Luss, R., Rigotti, M., Rish, I., Diachille, P., Gurev, V., Kingsbury, B., Tejwani, R., Bouneffouf, D.: Beyond backprop: Online alternating minimization with auxiliary variables. In Chaudhuri, K., Salakhutdinov, R., eds.: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA. Volume 97 of Proceedings of Machine Learning Research., PMLR (2019) 1193–1202

9. Riemer, M., Klinger, T., Bouneffouf, D., Franceschini, M.: Scalable recollections for continual lifelong learning. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press (2019) 1352–1359

10. Lin, B., Cecchi, G.A., Bouneffouf, D., Reinen, J., Rish, I.: A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry. In Seghrouchni, A.E.F., Sukthankar, G., An, B., Yorke-Smith, N., eds.: Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020, International Foundation for Autonomous Agents and Multiagent Systems (2020) 744–752

11. Lin, B., Bouneffouf, D., Cecchi, G.: Online learning in iterated prisoner's dilemma to mimic human behavior. arXiv preprint arXiv:2006.06580 (2020)

12. Lin, B., Cecchi, G., Bouneffouf, D., Reinen, J., Rish, I.: Unified models of human behavioral agents in bandits, contextual bandits and rl. arXiv preprint arXiv:2005.04544 (2020)

13. Noothigattu, R., Bouneffouf, D., Mattei, N., Chandra, R., Madan, P., Varshney, K.R., Campbell, M., Singh, M., Rossi, F.: Teaching AI agents ethical values using reinforcement learning and policy orchestration. In Kraus, S., ed.: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, ijcai.org (2019) 6377–6381

14. Bouneffouf, D., Rish, I.: A survey on practical applications of multi-armed and contextual bandits. CoRR **abs/1904.10040** (2019)

15. Lai, T.L., Robbins, H.: Asymptotically efficient adaptive allocation rules. Advances in Applied Mathematics **6**(1) (1985) 4–22

16. Bouneffouf, D., Parthasarathy, S., Samulowitz, H., Wistuba, M.: Optimal exploitation of clustering and history information in multi-armed bandit. In Kraus, S., ed.: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019, ijcai.org (2019) 2016–2022

17. Lin, B., Bouneffouf, D., Cecchi, G.A., Rish, I.: Contextual bandit with adaptive feature extraction. In Tong, H., Li, Z.J., Zhu, F., Yu, J., eds.: 2018 IEEE International Conference on

Data Mining Workshops, ICDM Workshops, Singapore, Singapore, November 17-20, 2018, IEEE (2018) 937–944

18. Balakrishnan, A., Bouneffouf, D., Mattei, N., Rossi, F.: Incorporating behavioral constraints in online AI systems. AAAI 2019 (2019)

19. Balakrishnan, A., Bouneffouf, D., Mattei, N., Rossi, F.: Using multi-armed bandits to learn ethical priorities for online AI systems. IBM Journal of Research and Development **63**(4/5) (2019) 1:1–1:13

20. Bouneffouf, D., Laroche, R., Urvoy, T., Feraud, R., Allesiardo, R.: Contextual bandit for active learning: Active thompson sampling. In: Neural Information Processing - 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part I. (2014) 405–412

21. Noothigattu, R., Bouneffouf, D., Mattei, N., Chandra, R., Madan, P., Varshney, K.R., Campbell, M., Singh, M., Rossi, F.: Interpretable multi-objective reinforcement learning through policy orchestration. CoRR **abs/1809.08343** (2018)

22. Balakrishnan, A., Bouneffouf, D., Mattei, N., Rossi, F.: Constrained decision-making and explanation of a recommendation (January 16 2020) US Patent App. 16/050,176.

23. Bouneffouf, D., Rish, I., Cecchi, G.A., Féraud, R.: Context attentive bandits: Contextual bandit with restricted context. In: IJCAI 2017, Melbourne, Australia, August 19-25, 2017. (2017) 1468–1475

24. Balakrishnan, A., Bouneffouf, D., Mattei, N., Rossi, F.: Using contextual bandits with behavioral constraints for constrained online movie recommendation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden. (2018) 5802–5804

25. Bouneffouf, D., Bouzeghoub, A., Gançarski, A.L.: A contextual-bandit algorithm for mobile context-aware recommender system. In Huang, T., Zeng, Z., Li, C., Leung, C., eds.: Neural Information Processing - 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III. Volume 7665 of Lecture Notes in Computer Science., Springer (2012) 324–331

26. Bouneffouf, D.: Exponentiated gradient exploration for active learning. Computers **5**(1) (2016) 1

27. Liu, S., Ram, P., Vijaykeerthy, D., Bouneffouf, D., Bramble, G., Samulowitz, H., Wang, D., Conn, A., Gray, A.G.: An ADMM based framework for automl pipeline configuration. In: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, AAAI Press (2020) 4892–4899

28. Allesiardo, R., Féraud, R., Bouneffouf, D.: A neural networks committee for the contextual bandit problem. In: Neural Information Processing - 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part I. (2014) 374–381

29. Bouneffouf, D., Claeys, E.: Hyper-parameter tuning for the contextual bandit. CoRR **abs/2005.02209** (2020)

30. Upadhyay, S., Agarwal, M., Bouneffouf, D., Khazaeni, Y.: A bandit approach to posterior dialog orchestration under a budget. CoRR **abs/1906.09384** (2019)

31. Bouneffouf, D.: Online learning with corrupted context: Corrupted contextual bandits. arXiv preprint arXiv:2006.15194 (2020)

32. Casella, G., Berger, R.: Statistical inference, duxbury advanced series in statistics and decision sciences. (2002)

33. Team, C.: R: A language and environment for statistical computing. R Foundation for Statistical Computing **3** (2013)

34. Tvedebrink, T.: dirmult: Estimation in dirichlet-multinomial distribution. R Package Version 0.1 **3** (2009)

35. Lesnoff, M., Lancelot, R.: aod: Analysis of overdispersed data. The Sage Development Team.(accessed 2011). http://www. sagemath. org **1** (2012)
36. Bishop, C.: Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn (2007)
37. Yee, T.W., et al.: The vgam package for categorical data analysis. Journal of Statistical Software **32**(10) (2010) 1–34
38. Stein, W.A., et al.: Sage mathematics software(version 5.0). R package version **1** (2012)